UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(4)

AD-A201 619

FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AIM 1061	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network		5. TYPE OF REPORT & PERIOD COVERED memorandum
		6. PERFORMING ORG. REPORT NUMBER
AUTHOR(s) Shimon Ullman and Amnon Sha'ashua		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0124
PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE July 1988
		13. NUMBER OF PAGES 15
MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution is unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) Unlimited		
18. SUPPLEMENTARY NOTES None		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) segmentation; image partitioning; saliency; visual perception; perceptual organization		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) see back of page		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 I

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

88 10 31 056

Abstract: When we look at images, certain salient structures often attract our immediate attention without requiring a systematic scan of the entire image. In subsequent stages, processing resources can be allocated preferentially to these salient structures. In many cases this saliency is a property of the structure as a whole, i.e., parts of the structure are not salient in isolation. In this paper we present a saliency measure based on curvature and curvature variation. The structures this measure emphasizes are also salient in human perception, and they often correspond to objects of interest in the image. We present a method for computing the saliency by a simple iterative scheme, using a uniform network of locally connected processing elements. The network uses an optimization approach to produce a "saliency map," which is a representation of the image emphasizing salient locations. The main properties of the network are: (i) the computations are simple and local, (ii) globally salient structures emerge with a small number of iterations, and (iii) as a by-product of the computations, contours are smoothed and gaps are filled in.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1061 July 1988

Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network

Shimon Ullman* and Amnon Sha'ashua+

Abstract: When we look at images, certain salient structures often attract our immediate attention without requiring a systematic scan of the entire image. In subsequent stages, processing resources can be allocated preferentially to these salient structures. In many cases this saliency is a property of the structure as a whole, i.e., parts of the structure are not salient in isolation. In this paper we present a saliency measure based on curvature and curvature variation. The structures this measure emphasizes are also salient in human perception, and they often correspond to objects of interest in the image. We present a method for computing the saliency by a simple iterative scheme, using a uniform network of locally connected processing elements. The network uses an optimization approach to produce a saliency map, which is a representation of the image emphasizing salient locations. The main properties of the network are: (i) the computations are simple and local, (ii) globally salient structures emerge with a small number of iterations, and (iii) as a by-product of the computations, contours are smoothed and gaps are filled in.

This report describes research partially done at the Massachusetts Institute of Technology within the Artificial Intelligence Laboratory. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124.

*Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139
+Dept. of Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

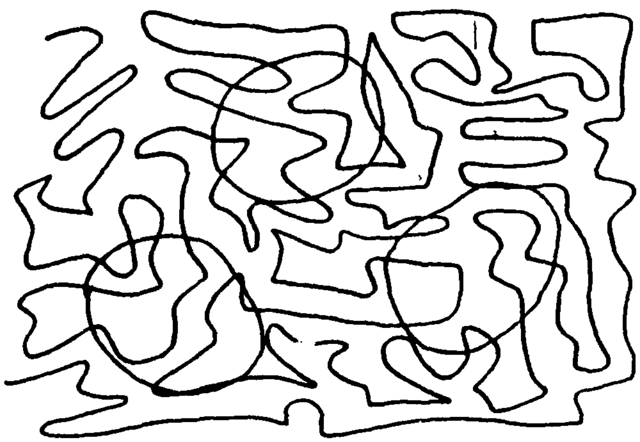
§1 Introduction

Salient structures can often be perceived in an image at a glance. They appear to attract our attention without the need to scan the entire image in a systematic manner, and without prior expectations regarding their shape. The processes involved in the perception of salient structures appear to play a useful role in segmentation and recognition, since they allow us to immediately concentrate on objects of interest in the image.

Consider the images in figures 1, 2 and 3. Certain objects in each image somehow attract our attention in a manner often described as 'preattentive'. For instance, the large blobs in Fig. *1a* and *1b* are prominent, although locally the blobs' contours are indistinguishable from background contours on the basis of local orientation, curvature, contrast, etc. It seems as if one must somehow capture most of the curve bounding a blob in order to perceive it as a prominent structure. The circle in Fig. 2 is immediately perceived although its contour is fragmented, implying that gaps do not hinder the immediate perception of such objects. In this case one must group together several line segments of the circle to distinguish it from the background. These examples also demonstrate that these prominent objects need not be recognized in order for them to be distinguished. The image in Fig. 3 is an edge image of a car in a cluttered background. Our attention is drawn immediately to the region of interest in the image. It seems that the car need not be recognized to attract our attention. When the image is inverted and presented for short periods, recognition becomes considerably more difficult, yet the same region remains salient.

The goal of this paper is to suggest what makes structures such as those in Fig. $1-3$ salient, and to propose a mechanism for detecting salient locations in an image. A locally connected network is proposed that can process images such as the figures above to construct a "saliency map", which is a representation of the image emphasizing salient locations. The computations of the net are devised to meet the following requirements: (i) the time it takes to detect a prominent structure does not depend on the complexity of background curves, (ii) curves may have any number of gaps, and (iii) the number of computations are restricted to the order of dozens or, at most, about a hundred steps in order to meet the time constraint involved in immediate perception.

Issues related to this problem include segmentation, perceptual organization, and figure/ground separation. Segmentation schemes have been investigated extensively in the field of computer vision and many algorithms have been suggested. They will not be reviewed here, since they are only marginally related to the problem at hand. Many of the segmentation processes that have been proposed were more ambitious than what is required, or what is possible, to achieve in the early stages where prominent areas are located. For example, they attempt to segment the entire image instead of just an area of interest. Our proposal is related to the suggestion made by Ullman (1986) that segmentation should be conducted on an area of interest rather than applied to the



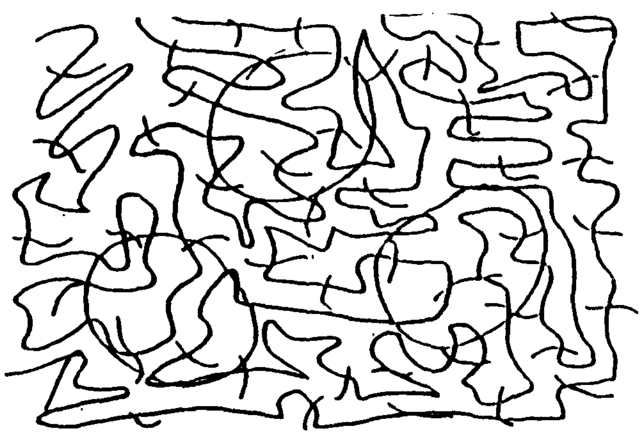


Figure 1a. Three prominent blobs are perceived immediately and with little effort. Locally, the blobs are similar to the background contours. (adopted from Mahoney (1986)

Figure 1b. Intersections were added to illustrate that the blobs are not distinguished by virtue of their intersections with the background curves.

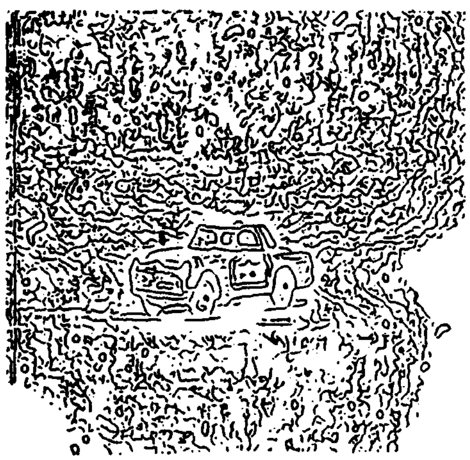


Figure 2. A circle in a background of 200 randomly placed and oriented segments The circle is still perceived immediately although its contour is fragmented.

Figure 3. An edge image of a car in a cluttered background. Our attention is drawn immediately to the region of interest. It seems that the car need not be recognized to attract our attention. The car also remains salient when parallel lines and small blobs are removed, and when the less textured region surrounding parts of the car is filled in with more texture.

entire image, implying that some preattentive process is required to detect prominent locations from which an area of interest is defined, prior to the act of segmentation.

Lowe's (1985) treatment of perceptual organization is more closely related to the problem addressed in this paper. The processes proposed by Lowe detect instances of collinearity, co-termination and parallelism among straight lines, and will not be effective in cases (e.g. Fig. 1) where these conditions do not play a major role. Most past approaches to segmentation also do not meet the requirements set above. In particular, they do not meet the time constraint and they depend critically on the complexity of the background curves.

1.1 Local and Global Saliency

The phenomena related to the perception of salient structures can be roughly divided into two classes. The first, referred to as *local saliency*, occurs when an element becomes conspicuous by having a simple distinguishing local property such as color, contrast, orientation, etc. For example, a red item placed among green ones immediately attracts attention by virtue of its unique color (Triesman and Galade 1980; Julesz 1981). The second case, referred to as *structural saliency*, occurs when the structure is perceived in a more global manner. That is, the local elements of the structure are not salient as in the former case but instead the arrangement of the elements is what makes the structure unique and salient.

We focus below on the saliency of curves, based on properties measured along them (the curves may be continuous or with any number of gaps). Not all phenomena of global immediate perception are necessarily accounted for by measuring properties of curves. For instance, one could measure the compactness of a structure, the degree of symmetry it contains and other measures that are region-based rather than curve-based. Nevertheless, properties of curves are often sufficient in order to separate objects from their background.

The fact that structural saliency requires measures that have a global extent introduces a severe complexity problem. The number of possible groupings of local line segments into curves, where the curves are allowed to have any number of gaps, explodes exponentially. The complexity issue becomes acute when considering the fact that a salient curve of a given length is not necessarily composed of salient sub-parts. Thus, contemporary pyramid techniques (see Rosenfeld 1986 for a review) would not be appropriate for detecting structural saliency, because they contain an implicit assumption that a salient curve is composed of salient sub-parts.

§2 Measuring Saliency as an Optimization Problem

Our goal is to construct a saliency map which is a representation of the image emphasizing salient locations. We seek to associate, therefore, a measure of saliency denoted by the function $\Phi(\cdot)$ to each location in the image. A property that seems to play a role in structural saliency is the combination of length and smoothness measured at a particular scale. That is, a measure of saliency that would account for the type of images above is one that favors long smooth curves, where the smoothness of a curve is related to its curvature or its curvature variation. We therefore face the following problems:

(1) Defining an appropriate measure Φ that, when applied to a point along a given curve, will increase when the curve increases in length and smoothness.

(2) A selection problem. The measure $\Phi(P)$ depends on the curve passing through P. Since the curves we are considering are either continuous or separated by any number of gaps, there will usually be many possible curves to consider. Our approach to this problem will be to select the curve that maximizes $\Phi(P)$ over all curves passing through P.

We defer the exact formulation of Φ until we have examined the manner by which it is computed. The reason is that the general method of computing Φ (using a simple local network) places strong constraints on the possible definition of Φ. In the next sections we describe the mechanism by which Φ is computed, and then derive an explicit formula for Φ.

2.1 The Basic Elements

We assume that Φ is computed by a locally connected network of processing elements. Our specific model is that at the level of computing saliency the image is represented by a network of $n \times n$ grid points, where each point represents a specific x, y location in the image. At each point P there are k *orientation elements* coming into P from neighboring points, and the same number of orientation elements leaving P to nearby points. Each orientation element p_i responds to an input image by signalling the presence of the corresponding line segment in the image, so that those elements that do not have an underlying line segment are associated with an empty area or gap in the image. We refer to a connected sequence of orientation elements $p_i, \dots, p_{i+N}$, each element representing a line-segment or a gap, as a curve of length N (note that curves may be continuous or with any number of gaps). The optimization problem is formulated as maximizing Φ_N over all curves of length N starting from p_i:

$$\max_{(p_{i+1},\dots,p_{i+N})\in\delta^N(p_i)} \Phi_N(p_i, \dots, p_{i+N})$$

where $\delta^N(p_i)$ is the set of all possible curves of length N starting from p_i.

A naive approach to this problem would involve an exhaustive enumeration of all combinations of $p_{i+1}, \ldots, p_{i+N}$ which would require an exponential search space of size k^N for each element in the network. In what follows, we will show that for a certain class of measures Φ ("extensible" measures), the computation becomes linear in N. We will then define a saliency measure Φ that measures length and smoothness, and at the same time is extensible and can be computed efficiently.

2.2 Multistage Optimization Approach

For a certain class of measures $\Phi(\cdot)$, the computation of Φ_N can be obtained by iterating a simple local computation. To illustrate, let us consider first curves that are only three elements long. The problem in this case is:

$$\max_{(p_{i+1}, p_{i+2}) \in \delta^2(p_i)} \Phi_2(p_i, p_{i+1}, p_{i+2})$$

That is, for a given element p_i, determine p_{i+1} (one of p_i's k neighbors) and p_{i+2} (a neighbor of p_{i+1}) such that $\Phi_2(p_i, p_{i+1}, p_{i+2})$ will be maximal. A naive approach will again require examining the k^2 different curves. Assume, however, that Φ_2 satisfies the condition:

$$\max_{\delta^2(p_i)} \Phi_2(p_i, p_{i+1}, p_{i+2}) = \max_{p_{i+1}} \Phi_1(p_i, \max_{p_{i+2}} \Phi_1(p_{i+1}, p_{i+2}))$$

In this case maximizing Φ_2 can be achieved by repeating the application of Φ_1 over shorter curves. The general approach is formulated in a similar manner:

$$\max_{\delta^N(p_i)} \Phi_N(p_i, \ldots, p_{i+N}) = \max_{p_{i+1} \in \delta(p_i)} \Phi_1(p_i, \max_{\delta^{N-1}(p_{i+1})} \Phi_{N-1}(p_{i+1}, \ldots, p_{i+N})) \qquad (2.1)$$

where $\delta(p_i)$ stands for $\delta^1(p_i)$. In this manner we reduce the search space needed for each curve of length N starting from p_i to the size of kN instead of k^N that is needed for the naive approach. The principle in (2.1) is related to the principle of optimality underlying all multistage decision processes, and in particular it is a special case of Dynamic Programming. We refer to the family of functions that obey the principle in (2.1) as *extensible functions*. We next derive an extensible function that prefers long curves that have low total curvature.

2.3 Deriving an extensible Function for Measuring Saliency

We next derive an expression for the saliency of an element p_i on a curve $\gamma = p_i, \ldots p_{i+N}$. γ is a curve of length N, terminating at p_i. (For a non-end element, the saliency is the sum of the contributions of the two sides.) Note that the saliency measure is associated with each element, not with the entire curve. Two elements along the same curve may have different saliency measures, depending on their position.

The saliency measure at p_i developed below has the general form $\sum_j \omega_{i,j}\sigma_j$, where j ranges over all the elements lying on γ. That is, the saliency at p_i is a weighted sum of the contributions from all the elements lying on the same curve.

Two factors play a role in the measure of saliency. The first factor is related to the length of the curve, and the second factor is related to its shape. The length of a curve is determined by the number of elements on the curve that have an actual curve (rather than a gap) passing through them. These elements are referred to as *active elements*, whereas the elements that are associated with gaps are referred to as *virtual elements*. To each element p_i we associate its local saliency σ_i. If p_i is an active element, then σ_i is set to be a positive value, which for the present is set to 1, and for a virtual element σ_i is set to 0. The measure related to the length of the curve $p_i, ..., p_{i+N}$ is:

$$\sum_{j=i}^{i+N} \sigma_j \tag{2.2}$$

The measure above is a sum of the local saliency values of the active elements along the curve. $\sum \sigma_j$ is in the range of 0 to $N+1$ depending on the number of active elements, implying that a continuous curve scores higher than a fragmented one of the same length. It is also possible to 'penalize' the existence of gaps, especially large ones, in order to attenuate the measure given to the curve when it is too fragmented. Penalizing the existence of gaps is obtained by associating an attenuation factor ρ_i with each element p_i. The value of ρ_i determines how quickly the contribution to the saliency from neighboring elements along the curve decays with distance. It is reasonable to use only to values for ρ_i, depending on whether p_i is an active or virtual element. If p_i is active then ρ_i is set to a value smaller or equal 1 (for the present it is set to 1). If p_i is virtual, then $\rho_i = \rho < 1$. We then define an attenuation function associated with the curve $p_i, ..., p_j$ as follows:

$$\rho_{i,j} = \prod_{k=i+1}^{j} \rho_k$$

where $\rho_{i,i} = 1$. The measure in (2.2) is modified by the attenuation factors:

$$\sum_{j=i}^{i+N} \rho_{i,j}\sigma_j \tag{2.3}$$

The measure in (2.3) is a weighted contribution of the local saliency values σ_j along the curve, where the weights are inversely related to the number of virtual elements along $p_i, ..., p_j$.

In order to measure the shape of the curve we use a measure that is inversely related to the total curvature of the curve. The total curvature of a curve γ is defined as $\int_\gamma \left(\frac{d\theta}{ds}\right)^2 ds$, where $\theta(s)$ is the slope along the curve, and $\frac{d\theta}{ds}$ at point P is known as the local curvature at that point (the inverse of R, the radius of curvature). We would like to use the total curvature to obtain a measure that is bounded, and is inversely

related to the total curvature. The following measure meets these requirements:

$$e^{-\int_\gamma (\frac{d\theta}{ds})^2 ds} \tag{2.4}$$

which is confined to values between 0 and 1. A straight line receives the value 1, and a meandering curve will approach the limit 0 as its total curvature grows to infinity. To obtain a discrete approximation to the measure in (2.4) we denote by α_k the orientation difference between the k'th element and its successor, and by Δs the length of an orientation element. The local curvature $\frac{1}{R}$ to the curve tangent to these elements (see Fig. 4) is:

$$\frac{2\tan\frac{\alpha_k}{2}}{\Delta s}$$

The arc's length is $\alpha_k R$, and therefore the total curvature square is approximated by:

$$\frac{2\alpha_k \tan\frac{\alpha_k}{2}}{\Delta s}$$

The discrete approximation to the total curvature measure along $p_i, ...p_j$ is therefore obtained by:

$$C_{i,j} = \prod_{k=i}^{j-1} f_{k,k+1}$$

where

$$f_{k,k+1} = e^{-\frac{2\alpha_k \tan\frac{\alpha_k}{2}}{\Delta s}} \tag{2.5}$$

$C_{i,j}$ plays the role of a weight given to each local saliency value σ_j along the curve. A measure that gives a high score to long curves with low total curvature is now defined as:

$$\sum_{j=i}^{i+N} C_{i,j}\rho_{i,j}\sigma_j \tag{2.6}$$

The measure in (2.6) is a weighted contribution of the local saliency values σ_j along the curve. Each weight is a product of two factors. The first factor is inversely related to the number of virtual elements along $p_i, ..., p_j$, and the second factor is inversely related to the total curvature of the curve. Curves that will receive a high measure on (2.6) are long curves that are as straight as possible and have the least number of gaps. The measure in (2.6) is also extensible according to the definition in (2.1) This can be shown by induction on the length of the curve, and the proof will not be detailed here.

Other functions for measuring the optimality of curves, using multistage optimization, were suggested by Ballard and Sklansky (1976), Martelli (1976) and Montanari (1971). The optimal curve in these cases is one that *maximizes* the sum of gray levels or edge magnitude along the curve, while minimizing the sum of orientation difference. In our terminology, the optimization function is:

$$\sum_{j=i}^{i+N} \sigma_j - \sum_{j=i}^{i+N} \alpha_j$$

This measure, however, is insensitive to the distribution of orientation difference along the curve and in general does not satisfy the requirement to prefer long and as-straight-as-possible curves.

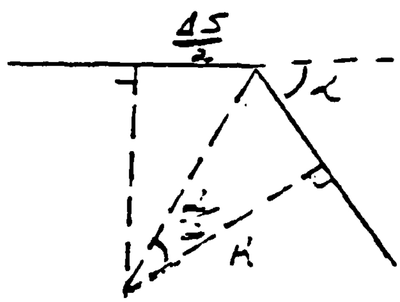


Figure 4. A discrete approximation to the curvature. R approximates the radius of curvature, α is the orientation difference, Δs is the length of both elements.

§3 The Saliency Network

In this section we summarize the computation performed by the network and its relation to the saliency measure defined above. The orientation elements constitute the basic computing elements of the net. Each element p_i is associated with a processor that can perform some computation based on its state and the state of its k neighboring processors. This defines a uniform network containing kn^2 processing units, with local communication. In the current implementation k is equal to 16, providing a reasonable angular resolution.

3.1 Computation of Elements in the Network

With each element p_i is associated a state variable denoted by E_{p_i} and a set of three attributes that includes its local saliency σ_i, its orientation θ_i and its attenuation

factor ρ_i. Each element p_i updates its state variable E_{p_i} iteratively through a local computation. (We use here the notation E_{p_i} to indicate explicitly that the variable is associated with the element $p-i$ in the network.) At the end of iteration N, E_{p_i} contains the measure of saliency derived in (2.6) which will be maximal over all possible curves of length N starting at p_i, where these curves are either continuous or with any number of gaps.

E_{p_i} is updated by the following computation:

$$E_{p_i}^{(0)} = \sigma_i$$

$$E_{p_i}^{(n+1)} = \sigma_i + \rho_i \max_{p_j \in \delta(p_i)} E_{p_j}^{(n)} f_{i,j} \tag{3.1}$$

where p_j is one of k possible neighbors of p_i, and $f_{i,j}$ are the "coupling constants" defined in (2.5). To unravel the recurrence formula above, we isolate a specified curve γ represented by $\theta_i, ..., \theta_{i+N}$ where each element along the curve has only a single neighboring element to communicate with. The following proposition relates the value of the state variable of p_i with the measure in (2.6).

Proposition 1:

$$E_{p_i}^{(N)} = \sum_{j=i}^{i+N} C_{i,j}\rho_{i,j}\sigma_j$$

The proof is by induction on the length of the curve and will not be detailed here. The proposition above together with the fact that the measure is extensible implies that among all possible curves γ_i of length N starting from p_i, either continuous or with any number of gaps, E_{p_i} will be computed along that curve which is maximal with respect to the measure in (2.6), namely

$$E_{p_i}^{(N)} = \max_{\gamma_i} \sum_j C_{i,j}\rho_{i,j}\sigma_j$$

taken over all γ_i. It is worth noting that the fact that the measure Φ is extensible, does not imply that the optimal contour through P simply extends itself as the iterations proceed. In fact, the optimal curve at stage $N+1$ can be different from the optimal curve at stage N

The state values of elements in the network form a new representation of the image which is a 'biased' view of the visual environment, emphasizing interesting or conspicuous locations. We denote this representation as the *saliency map.* The term of saliency map was used by Koch and Ullman (1986) for representing (using our terms) local saliency.

3.2 Additional Properties of the Network

Convergence of the State Values

The concept of an iterative computation raises the issue of convergence when the number of iterations goes to infinity. This issue is important in the context of the saliency network because an element p_i might be influenced by its own state in a feedback loop if it lies on a closed curve. The following proposition considers a closed curve and evaluates the state of an element of the curve after an infinite number of iterations.

Proposition 2:

Consider $p_i, ..., p_{i+N}$ a closed curve where $p_i = p_{i+N+1}$. The state of p_i converges to the following value:

$$E_{p_i}^{(kN)} \underset{k\to\infty}{\longrightarrow} \frac{E_{p_i}^{(N)}}{1 - C_{i,i+N}}$$

The proof is by induction on the length of the curve. The main point to notice is that a closed curve (even if it is fragmented) will increase its value when the number of iterations exceeds the curve's perimeter. If we consider a continuous circle of radius r, for example, then $C_{i,i+N} = e^{-\frac{2\pi}{r}}$ which is always less than 1. In practice, the increase is considerably smaller than the limiting value because we perform a restricted number of iterations.

Tracing the Curve Starting From a Given Element

The computation performed by each element includes a local preference between neighboring elements. That is, at each iteration each element p_i selects the neighbor p_j that contributes the most to its state. The information regarding local preference can be used to trace a linked curve starting from p_i in a recursive manner, namely, p_j is the second element in the curve, p_j's preferred neighbor is the third element, etc. Given a conspicuous element as a starting point, we could extract the curve that is optimal according to (2.6). Examples of these curves are given in section 4.

Filling Gaps by the Saliency Network

The ability to cope with gaps is important for the applicability of the saliency network to real images. Edge maps obtained from real images are often corrupted by multiple gaps, and what seems as a smooth salient curve often turns out to be fragmented after edge detection has been applied.

A virtual element (that lies in a gap) participates in the computation of (3.1) in a similar manner to active elements. Consider for instance a gap starting from p_{j+1} and ending at p_{j+k}. That is, p_j is an active element, but $p_{j+1}, ..., p_{j+k}$ are virtual elements. An element will update its state provided that it has at least one neighbor with a state value different from 0. It will take at most k iterations for p_{j+k} to update its state.

The network will fill-in a curve γ_i that will maximize the value of $\rho^{|\gamma_i|}C_{j,j+k}$. That is, the preference is for filled-in curves having low total curvature $C_{j,j+k}$, while minimizing their overall length $|\gamma_i|$. The relative weight of the two factors is controlled by setting the values of ρ. In the current implementation ρ was set to 0.7, which was found experimentally to give results that are generally in agreement with our own perception. The curves generated in this manner are similar (for orientation difference less than $\frac{\pi}{2}$) to several other methods for completing gaps in contours and for modelling subjective contours in human perception (Rutkowski 1979; Ullman 1976; Webb and Pervin 1984).

3.3 Additional Computations Performed by the Network

Measure of Saliency Based on Low Curvature Variation

The computation of the network summarized in (3.1) produce a saliency map based on the measure in (2.6). This does not rule out the possibility of additional properties that mediate structural saliency. For instance, the blobs in Fig. 1 seem to be prominent on the basis of low curvature variation rather than low overall curvature. A second saliency measure was therefore formulated that prefers long curves with low total curvature variation. Details of this second measure can be found in (Sha'ashua 1988). As a result, the saliency network constructs two saliency maps, one for each property, from which salient locations can be detected.

Smoothing the Measured Curves

The input to the saliency network is an edge map that determines which of the network's elements are active. The edges in the edge map are often noisy, due to sensor noise, quantization effects, and various effects of the edge detection process. Reducing noise is important because what appears to be a smooth curve to our visual system may turn out to be rather serrated at the edge map level. Smoothing can be obtained in part by analyzing the same image at different resolutions. It turns out, however, that some smoothing is often desired within a given scale of analysis.

A naive approach would be to extract all curves, replace them by a smooth approximation and then apply the saliency network to the smoothed curves. However, such an approach will encounter the same complexity issue regarding the number of possible curves discussed in section 1.1. We handle the problem of smoothing curves as a local computation that is performed within the saliency network itself, as an integral part of computing the saliency measure. In a nutshell, the coordinates associated with each orientation element are modified in an iterative manner, to smooth the curve passing through that element. The approach underlying the computation is to associate an en-

ergy level to each curve so that the smooth approximation is of minimum energy. The energy functional is given by:

$$E(\gamma) = \frac{1}{2}\lambda \sum_{j=i}^{i+N} \left((x_j - x_j^{(0)})^2 + (y_j - y_j^{(0)})^2\right) \; + \; \frac{1}{2}\int_\gamma \left(\frac{d^2\theta}{ds^2}\right)^2 ds$$

where (x_j, y_j) $j = i, ..., i + N$ are the coordinates of the smooth approximation to the curve (x_j^0, y_j^0) $j = i, ..., i + N$. A curve of minimum energy is one that minimizes its total curvature variation while being as close as possible to the original curve. The parameter λ controls the relative weight between the two terms (for a similar energy functional see Poggio *et al.* (1985)). The energy is lowered at each iteration in a process that involves only local computations. These local computations are combined with those in (3.1), resulting a network which measures saliency of curves while smoothing them simultaneously. The details can be found in (Sha'ashua 1988).

§4 Examples of the Saliency Network at Work

The main issues illustrated by the examples are (i) the saliency map, and (ii) the by-product creation of linked curves, which is a by-product of the saliency computation.

Prominent locations in the image are represented as elements having a high measure of saliency as computed by the network. For illustration purposes the saliency map will be displayed as a gray-level image in which an element p_i is displayed as a bar of width ω_i and intensity value τ_i, given by:

$$\tau_i = \left\lfloor \frac{E_{p_i}}{\max_{p_j} E_{p_j}} 255 \right\rfloor \qquad \omega_i = \left\lceil \frac{\tau_i}{255} 4 \right\rceil$$

In other words, increased saliency measure corresponds to an increase in brightness and in width of the element in the display. The most salient element is displayed as a white bar of width four, and the least salient element is displayed as a black segment.

The first example is a synthetic image (not produced by edge detection) shown in Fig. 2. It is constructed from a fragmented circle placed among a background of randomly placed and oriented elements. The number of background elements is 200 and the circle consists of 60 elements. The circle is immediately perceived by our visual system. The saliency network is applied to this image for ten iterations. Fig. 5 presents the saliency map after that period, and Fig. 6 presents the selected curve starting from the most salient element. The result is in agreement with the perception of the circle by our visual system. The saliency measure of each element of the circle is significantly higher than the measure given to the background elements. In this regard, the circle virtually 'pops-out' from the saliency map.

The second point to notice is that a complete object is separated from the background although it is initially fragmented. This agrees with the observation that perception is not severely affected by the presence of gaps. The final point to notice is

that although the length of the salient curve is 60 elements, the number of iterations required for distinguishing the circle from its background is considerably smaller. This happens because although each element of the circle is not salient by itself, groups of ten elements already become sufficiently salient. Outside the circle, the probability of having a low curvature chain of length ten is low. In fact, the probability remains small even when the number of background elements increases considerably. To illustrate, we doubled the number of background elements as shown in Fig. 7. We applied again ten iterations to produce the saliency map in Fig. 8. Starting from the most salient element, the curve extracted by the network is identical with the one in Fig. 6.

The next example is the image in Fig. 3. Fig. 9 shows the saliency map after 30 iterations. Only the region surrounding the car is displayed. The saliency measure given to most of the elements of the car is significantly higher than that given to the background elements. Fig. 10 displays the five most salient curves obtained by tracing the most salient elements. Note that the traced curves have been smoothed, and that the gaps have been filled in. The results suggest that the saliency computation is useful for distinguishing significant structures in the image.

The final example is the image in Fig. 1*a*. The input to the network was obtained by edge detection from the original hand-drawn image. We show the results for a part of the image containing one of the blobs. Fig. 11 displays the saliency map for low curvature variation after 160 iterations, which is twice the number of elements on the perimeter of the blob. The elements of the blob become stronger than the background elements after 70 iterations, in agreement with the observation that one must capture almost the entire blob in order to perceive it as prominent. Interestingly, the results of the low curvature map are similar, but about 100 iterations are required for the blob to become prominent. Fig. 12 displays the curve starting from the most salient element. In this case also the curve is smoothed by the network while measuring its saliency.

§5 Summary

5.1 Brief Summary of the Scheme

A measure of saliency $S(P)$ is defined for the edge elements in the image. The saliency measure is used for detecting globally salient structures in the image. As a by-product, the process fills-in smoothly gaps in fragmented contours, and provides linking information between edge segments.

Saliency of a Single Curve

Let γ be a curve and P an end element of the curve. The saliency of P given γ,

$S_\gamma(P)$ is defined as:

$$S_\gamma(P) = \sum_i \omega_i \sigma_i$$

where σ_i is the local saliency of the i'th edge element along γ, and ω_i is the weight of the element's contribution.

The weight ω_i is a product of two factors. The first factor is

$$e^{-c_i}$$

where c_i is the total curvature of the curve up to the i'th element. The second factor penalizes the existence of gaps, and is defined as:

$$\prod_{k=0}^{i} \rho_k$$

where ρ_k is the attenuation factor of the k'th element along the curve. One value is used for real edges, another for gaps in the contour.

The Saliency Measure

The measure in section 1 depends on a particular curve γ. The saliency at P is given by:

$$S(P) = \max_\gamma S_\gamma(P)$$

A maximum is reached over *all* possible curves terminating at P. In practice, the definition is limited to curves of length N:

$$S_N(P) = \max_{\gamma_N} S_{\gamma_N}(P)$$

Note: the maximum is taken over all possible curves, including fragmented ones. As a by-product, curves are being filled-in.

Operation of the Network

$$E_i^{(0)} = \sigma_i$$

$$E_i^{(n+1)} = \sigma_i + \rho_i \max_{j \in \delta(i)} E_j^{(n)} f_{i,j}$$

$\delta(i)$ are all the neighbors of element i. The quantities ρ_i and $f_{i,j}$ ("couplings") are constants of the network. The initial input at step 0 are the local saliencies σ_i. At the $(n+1)$ iteration, each element simply adds the maximal contribution from its neighbors to its own local saliency. After N iterations the computation defined above computes the saliency measure $S_N(P)$ for each element P.

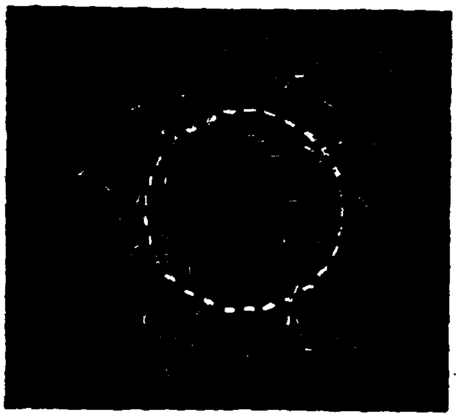
5.2 General Summary

It is proposed that immediate perception includes processes for detecting salient structures in the image on which subsequent processes such as segmentation and recognition can focus. The saliency of a structure is divided into two sources, local saliency,

and structural saliency. Of the two, structural saliency is more problematic from a computational point of view since it requires the efficient computation of certain global properties.

A locally connected network was devised to produce a saliency map, which is a representation of the image emphasizing salient locations. The network exhibits the following properties: (i) the computations are local and simple, (ii) the number of computations are in the order of dozens or up to about a hundred, (iii) there is little dependence on the complexity of the image, (iv) gaps in curves are filled in the course of the computation, (v) contours are *smoothed* in the course of producing a saliency map, (vi) the network produces linking information so that curve tracing across junctions, branches and gaps is possible, and (vii) the network is robust in the sense that malfunction of some processing units does not affect seriously the performance of the network.

Acknowledgement: We thank E. Grimson for his comments.



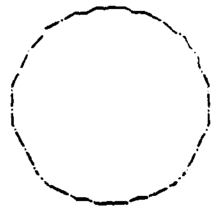


Figure 5. Saliency map of the image in Fig. 2 obtained by the network after 10 iterations. The saliency measure of each element of the circle is significantly higher than of the background elements.

Figure 6. The curve starting from the strongest element in figure 5. Virtual elements are displayed as dotted lines.

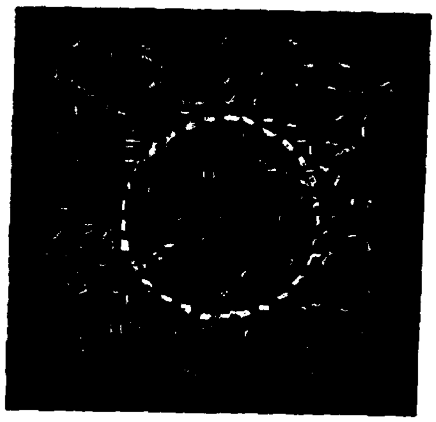
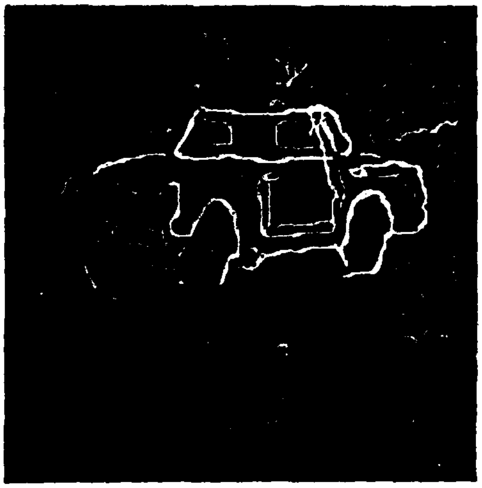


Figure 7. The same circle as in figure 2 but with 400 background segments.

Figure 8. Saliency map of the image in Fig. 7 obtained by the network after 10 iterations.



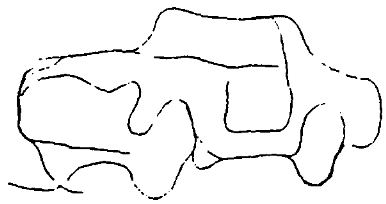
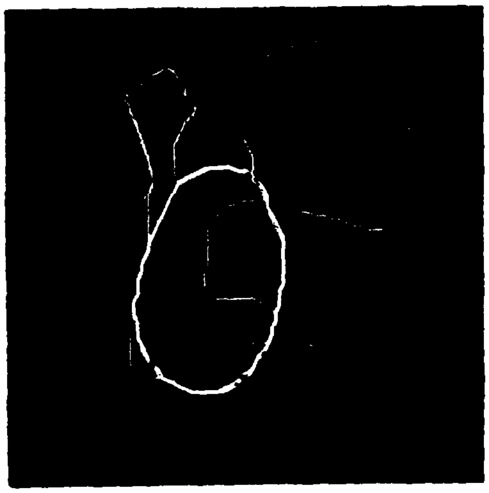


Figure 9. Saliency map of the image in Fig. 3 obtained by the network after 30 iterations. The region of interest virtually 'pops-out' from the display.

Figure 10. The five most salient curves obtained by tracing the most salient elements of figure 9. The curves have been smoothed and gaps have been filled in.



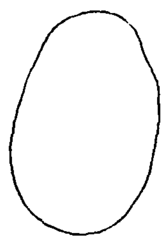


Figure 11. Saliency map for low curvature variation of the image in Fig. 1

Figure 12. The curve starting from the strongest element in figure 11 is traced. The curve is smoothed by the network while measuring its saliency.

References

Ballard,D.H and Sklansky,J (1976), *A ladder-structured decision tree for recognizing tumors in chest radiographs.* IEEE Trans. Comput., Vol C-25, 503-513

Julesz,B (1981), *Textons, the elements of texture perception, and their interactions.* Nature, 290, 91-97

Koch,C and Ullman,S (1986), *Shifts in selective visual attention: towards the underlying neural circuitry.* Bat Sheva seminar on selective attention 165-173 Jerusalem, 1986

Lowe (1985), ***Perceptual Organization and Visual Recognition.*** Boston: Kluwer Academic Publishers.

Mahoney,J.V (1986), *Image chunking: defining spatial building blocks for scene analysis.* S.M. thesis, Dept. of electrical engineering and computer science, M.I.T.

Martelli,A (1976), *An application of Heuristic search methods to edge and contour detection.* Comm. Assoc. Comput. Mach., Vol 19, 73-83

Montanari,U (1971), *On the optimal detection of curves in noisy pictures.* Comm. Assoc. Comput. Mach, Vol 14, 335-345

Poggio,T, Voorhees,H and Yuille,A (1985), *A regularized solution to edge detection.* MIT AI memo 833, May 1985.

Rosenfeld,A (1986), *Some pyramid techniques for image segmentation.* CS-TR-1664 University of Maryland

Rutkowski,W.S (1979), *Shape completion.* Computer Graph. Image Proc.,9, 89-101

Sha'ashua,A (1988), *Immediate perception of global structures: A computational approach.* M.Sc thesis, Dept. of Applied Math., Weizmann Institute of science

Triesman,A and Galade.G (1980), *A feature integration theory of attention.* Cog. Psychol., 12, 97-136

Ullman,S (1976), *Filling in the gaps: The shape of subjective contours and a model for their generation.* Biol. Cyb., 25, 1-6

Ullman,S (1986), *An Approach to Object Recognition: Aligning Pictorial Descriptions.* MIT AI memo 931, December 1986.

Webb,J.A and Pervin,E (1984), *The shape of subjective contours.* Proc. IAAA-84 340-343